Overview

- Modeling of fMRI Signal
- Modeling of fMRI Noise

Signal: Examples of fMRI Data

- Strong Signal
- Moderate Signal
- Weak Signal

Signal: Components

- BOLD Effect
  - ⇀ Related to experimental paradigm
  - ⇀ Temporal delay and blurring
- Physiological Effects
  - ⇀ Cardiac & Respiration
- Movement
  - ⇀ Edge Artifacts
- Drift
  - ⇀ Aliased physiological effects
  - ⇀ Scanner instability

Signal: Components - Examples

- BOLD Effect
- Physiological Effects
  - ⇀ Veins/Arteries
- Movement
  - ⇀ Rim artifacts
- Drift

Signal: Modeling Components

- BOLD Effect
  - ⇀ In GLM
- Physiological Effects
  - ⇀ Prospectively
  - ⇀ Possibly collecting pulse and respiration
- Movement
  - ⇀ Prospectively
- Drift
  - ⇀ In GLM

Signal: BOLD Modeling w/ GLM

- GLM requires predictors
- How to define?
  - ⇀ Box-car ____________________
  - ⇀ Shifted Box-car ____________________
  - ⇀ Draw by hand? ____________________
  - ⇀ Linear system ____________________

Signal: BOLD Modeling w/ LTI

- Linear Systems Theory
  Neuronal Resp. ⇒ Hemodynamics ⇒ MRI Physics ⇒ Obs
  - ⇀ Totally specified by
    - – Driving input
    - – Impulse response function (IRF)
- Easy to specify
  - ⇀ Convolve IRF with impulse
  - ⇀ "Convolution" like shift-then-summing a waveform

Signal: BOLD Modeling w/ LTI What IRF?

- We don't know the exact form of IRF
- Estimate Parameterized IRF
  - ⇀

D:CRGhrf

  - ⇀ Not a linear modeling problem
- Estimate an arbitrary IRF
  - ⇀

D:Deconv

  - ⇀ Linear, but lots of parameters

Signal: BOLD Modeling w/ LTI What IRF?

- Assume it
  - ⇀ Typical solution
  - ⇀ Relies on validity of "canonical" HRF
- Estimate an arbitrary IRF once, assume it subsequently
  - ⇀ For each subject, perform simple visual or motor task
  - ⇀ Use that data to *deconvolve* IRF
  - ⇀ Use that IRF for rest of subject's data
  - ⇀ Assumes IRF consistent across the brain

Signal: BOLD Modeling w/ LTI Examples

- Event Related
- Block Design

Signal: BOLD Modeling w/ LTI Shortcomings

- Assumes BOLD is a linear time-invariant system
  - ⇀ 20 quickly spaced events will produce predicted response almost 20-times as high as single event
  - ⇀ Can't account for nonlinear interactions between events
    - – For example, refractory period following an event
  - ⇀ Seems to work *OK* for events as close as 2s
    - – But really not linear, even for 20s ISI (Vazquez & Noll)
- Antisymmetry
  - ⇀ Convolution implies rise is antisymmetric to fall
  - ⇀ Does this make sense?

Signal: Linear Modeling Tricks

- Cannot parameterize shift with a linear model
- Can approximately capture shift with a linear model
  - ⇀ Predictor: $g(t)$
  - ⇀ Shifted predictor: $g(t+\delta)$
  - ⇀ Taylor series approximation about $\delta = 0$

$$g(t+\delta) \approx g(t) + g'(t)\delta$$

  - ⇀ Hence including a derivative accounts for small shifts
- Example

Signal: Basis of IRFs

- Choosing a single IRF can be restrictive
- More general is to use family or *basis* of IRF's
  - ⇀ Any collection of smooth functions works
- More flexible, but then can't use a t test
  - ⇀ Have to use F test to assess all elements
  - ⇀ Difficult to compare conditions

Signal: Modeling Drift

- Drift is observed in phantoms, cadavers, and grad students
  - ⇀ A very significant source of variation
- Drift has no precise definition
  - ⇀ Any slow variation
- Typical modeling approach is to use Discrete Cosine Basis
  - ⇀ Example
- Other bases may be better

Signal: Modeling Drift

- Must take care to distinguish drift from experimental signal
- If not careful, can "suck" signal
  - ⇀ If drift basis looks like BOLD predictor BOLD effect won't be significant
  - ⇀ Usual rule is to set lowest period to twice experimental period

N:cfDCT

Signal: BOLD + Drift Example

- Subject presented with quickly flashed checkerboards

Signal: Quiz

I-1 In a linear system, only two things are needed to define the output. What are they?

I-2 Identify & explain a dodgy assumption of linear system approach.

II-3 What's the danger in not modeling drift at all?

II-4 What's the danger in modeling too much drift (too much, i.e. using too many cosines)?

Noise: Overview of fMRI Noise

- fMRI noise is not independent
  - ⇀ Exhibits temporal autocorrelation
    - – Physiological effects
    - – Scanner instability artifacts
- Null fMRI Data
- Independent data

Noise: Three Approaches

- Independence
  - ⇀ Do nothing
- Prewhitening
  - ⇀ Do the right, difficult, not-so-robust thing
- Precoloring
  - ⇀ Do the somewhat right, easier, robust thing

Noise: Independence Approach

- Independence
  - ⇀ Ignore autocorrelation
  - ⇀ Dangerous!
    - – $\hat{\sigma}^2$ is underestimated
    - – Significance overestimated
    - – p-values too small

Noise: Prewhitening Approach

- Statistically optimal approach is to decorrelated the data
- Correlated model
  - ⇀ $Y = X\beta + \epsilon$
  - ⇀ $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 V)$, $V \neq I_n$
  - ⇀ V is correlation matrix
- Decorrelate model
  - ⇀ $V^{-1/2}$ is matrix s.t. $V^{-1/2}V^{-1/2} = V^{-1}$
  - ⇀ Premultiply model by $V^{-1/2}$
  - ⇀ $V^{-1/2}Y = V^{-1/2}X\beta + V^{-1/2}\epsilon$

$$\begin{aligned}\text{Var}(V^{-1/2}\epsilon) &= V^{-1/2}\text{Var}(\epsilon)V^{-1/2}\\ &= V^{-1/2}(\sigma^2 V)V^{-1/2}\\ &= \sigma^2 V^{-1/2}(V)V^{-1/2} = \sigma^2 I_n\end{aligned}$$

Noise: Prewhitening Approach

- Decorrelating works!
  - ⇀ $V^{-1/2}\epsilon$ are independent, or “white”
- Can now apply usual GLM to
  - ⇀ Data: $V^{-1/2}Y$
  - ⇀ Model: $V^{-1/2}X$
  - ⇀ Interpretation of parameters uneffected
- But what is V?
  - ⇀ We don’t know it, we have to estimate it: $\hat{V}$
  - ⇀ But above math is for V not $\hat{V}$
- The randomness in $\hat{V}$ will corrupt inferences on β

Noise: Precoloring Approach

- Assume that autocorrelation is mild
- Assume that is OK to temporally smooth
  - ⇀ Not unreasonable, since BOLD response temporally blurs
- But smoothing induces autocorrelation!
  - ⇀ Exactly!
  - ⇀ We want the smoothing to swamp the intrinsic autocorrelation
  - ⇀ Because then we know the *exact* form of autocorrelation

Noise: Precoloring Approach

- Precoloring Steps
  - ⇀ Smooth data & model matrix
  - ⇀ Assume that *any* autocorrelation is due to smoothing
  - ⇀ Fit GLM w/ usual $\hat{\beta}$
  - ⇀ Use modified std errors
  - ⇀ Compute t- & F- statistics as usual
  - ⇀ Compute adjusted degrees of freedom for t & F p-values

Noise: Precoloring Approach Details

- Smooth model
  - ⇀ Let K by a $n \times n$ smoothing matrix
  - ⇀ Premultiply model by K
    - – $KY = KX\beta + K\epsilon$
  - ⇀ And (ignorantly) apply standard GLM estimate
    - – $\hat{\beta} = ((KX)'(KX))^{-1}(KX)'KY$
- $\hat{\beta}$ is unbiased
  - ⇀ $\mathsf{E}(\hat{\beta}) = \beta$
  - ⇀ (But not optimal: $\mathsf{Var}(\hat{\beta})$ would be smaller w/ prewhitening)

Noise: Precoloring Approach Details

- But standard variance result is wrong
  - ⇀ $\mathsf{Var}(\hat{\beta}) \neq ((KX)'(KX))^{-1}\sigma^2$
  - ⇀ Correct result is a mess

$$\begin{aligned}
\mathsf{Var}(\hat{\beta}) &= ((KX)'(KX))^{-1}(KX)'K\mathsf{Var}(\epsilon)K'(KX)((KX)'(KX))^{-1}\sigma^2 \\
&= ((KX)'(KX))^{-1}(KX)'K \quad V \quad K'(KX)((KX)'(KX))^{-1}\sigma^2 \\
&\approx ((KX)'(KX))^{-1}(KX)'K \quad I_n \quad K'(KX)((KX)'(KX))^{-1}\sigma^2
\end{aligned}$$

- – Approximation comes from ignoring intrinsic autocorrelation

Noise: Precoloring Approach Details

- Degrees of freedom
  - ⇀ Define residual forming matrix R
    - – $R = I_n((KX)'(KX))^{-1}(KX)'$
  - ⇀ Then degrees of freedom are
  - ⇀ $\nu = \text{trace}(RV)^2/\text{trace}(RVRV)$
    - – trace() is the sum of the diagonal

Noise: Precoloring Approach Details

- This approach equivalent to Greenhouse-Geisser
  - ⇀ However, we don't use a conservative lower bound
  - ⇀ We assume a autocorrelation so we can work out df exactly
- Prewhitening was optimal. Is precoloring bad?
  - ⇀ Optimality of prewhitening assumes V known
  - ⇀ If $\hat{V}$ used, have to do further study
- Friston "To Smooth, or Not to Smooth"
  - ⇀ Prewhitening has less variable $\hat{\beta}$'s, but $\hat{\sigma}^2$ off
  - ⇀ Precoloring has more variable $\hat{\beta}$'s, but more accurate $\hat{\sigma}^2$

Noise: Precoloring Approach Conclusion

- For now, temporal smoothing is pragmatic solution
  - ⇀ When we can get robust, computationally efficient estimates of autocorrelation ($\hat{V}$), prewhitening will become defacto

Noise: Autoregressive Models - AR(1)

- Autoregressive model for errors
  - ⇀ Each error includes a fraction of the last
  - ⇀ $\epsilon_i = \rho\epsilon_{i-1} + \eta_i$
    - – ρ is AR parameter
    - – η_i is an independent perturbation
- AR(p) Models consider more of past
  - ⇀ $\epsilon_i = \rho_1\epsilon_{i-1} + \rho_2\epsilon_{i-2} + \cdots + \rho_p\epsilon_{i-p} + \eta_i$
- AR models are special case of prewhitening
- They are just a specific model form for $V = \mathsf{Var}(\epsilon)$

Noise: BOLD + Drift Example w/ Smoothing

- Subject presented with quickly flashed checkerboards

Noise: Summary of Autocorrelation Modeling

Model	Autocorrelation V known or estimated?	Estimation of β	Variance of $\hat{\beta}$	Estimation of σ^2
Independent	n/a	Unbiased	Suboptimal	Biased-worst
Prewhitening	Known	Unbiased	Optimal	Unbiased
Prewhitening	Est.	Unbiased	Suboptimal*	Biased-worse*
Precoloring	Known	Unbiased	Suboptimal	Unbiased
Precoloring	Est.	Unbiased	Suboptimal-worse*	Biased*

* From "To Smooth or Not to Smooth"

K.J. Friston, O. Josephs, E. Zarahn, A.P. Holmes, S. Rouqette, and J.-B. Poline. "To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis." NeuroImage, 12:196-208, 2000.

Noise: Summary of Autocorrelation Modeling

- Friston *et al.* temporal smoothing justification
  - Accept suboptimal precison for $\hat{\beta}$,
  - In exchange for better precision of $\hat{\sigma}^2$, *and*
  - Robustness with respect to wrong intrinsic autocorrelation
    - With smoothing, get same results assuming either
      $\text{Var}(\epsilon) = I_n$
      $\text{Var}(\epsilon) = \hat{V}$ (e.g. from AR(p) model)
- Note this is *intra*subject modeling issue
  - If only interest is in *inter*subject, random effects inference, none of this matters
  - In that case, only need unbiased $\hat{\beta}$, which the independence model offers

Noise: Quiz

II-1 What is the principal danger in using a model of independence on fMRI data?

II-2 The whitening approach is optimal, in that it produces minimum variance unbiased estimators. Why doesn't everyone use it?

II-3 In the precoloring approach, by smoothing we are throwing away information (the "high resolution" information). Why is this not so much a problem?

II-4 Both the prewhitening and precoloring approach involve premultiplying the model by a square $n \times n$ matrix. $V^{-1/2}$ in one case, K in the other. What is the key difference between these two matrices?